Chapter 3: Statistical Description of Data

El Mechry, El Koudouss

Fordham University

February 19, 2021

Introduction

In this chapter we will cover:

- measures of Central Tendency
  - Arithmetic mean
  - Weighted mean
  - Median
  - Mode
- Measures of Dispersion
  - Range
  - Quantiles
  - Mean absolute deviation
  - Variance
  - Standard deviation

A Reminder

Remember the difference between the terms: *Population* and *Sample*.

Our goal is to use information about a sample to make inferences about the population from which the sample was drawn.

Characteristics of the population are referred to as *parameters*, while characteristics of the sample are referred to as *statistics*.

We will be using sample statistics to predict population parameters.

Measures of Central Tendency: The Arithmetic Mean

- The arithmetic mean (Also known as arithmetic average, or average) is sum of the data values divided by the number of observations.
- We denote the sample mean by $\bar{x}$ and population mean by μ (pronounced "myew").
- The arithmetic mean can be expressed as

$$\bar{x} = \frac{\sum\limits_{i=1}^{n} x_i}{n}$$

- Notice this expression is also true for population mean, with the sample size n replaced by population size N.

The Arithmetic Mean: An Example

The following table shows net worth data of major 2016 election candidates.

Candidate	Net worth (Millions)
Sanders	0.7
Rand Paul	2
Christie	3
Cruz	3.5
Huckabee	9
Kasich	10
Jeb Bush	22
Carson	26
Chafee	32
Hillary Clinton	45
Fiorina	58
Trump	4500

The Arithmetic Mean: An Example

Calculating the average net worth of 2016 election candidates:

Start with the observations: $x_1 = 0.7$, $x_2 = 2$, $x_3 = 3$, $x_4 = 3.5$, ...

Expand the sum: $\sum_{i=1}^{n} x_i = \sum_{i=1}^{12} x_i = x_1 + x_2 + ... + x_{12}$

Add up the observations:
$$\sum_{i=1}^{12} x_i = 0.7 + 2 + 3 + 3.5 + 9 + 10 + 22 + 26 + 32 + 45 + 58 + 4500 = 4711.2$$

Calculate the mean: $\bar{x} = \sum_{i=1}^{12} /12 = 4711.2/12 = 392.6$

So, the average net worth of a 2016 presidential candidate is \$392.6 Million.

This data shows that outliers can distort the Arithmetic Mean. Notice that without Trump the average newt worth falls from \$392.6 to \$19.2 Million.

The Arithmetic Mean: Example

The following is a list of closing stock prices:

31.69, 56.69, 65.50, 83.50, 56.88, 72.06, 121.44, 97.00, 42.25, 71.88, 70.63, 35.81, 83.19, 43.63, 40.06.

Calculate the average closing price.

Answer:

The Arithmetic Mean: Example

The following is a list of closing stock prices:

31.69, 56.69, 65.50, 83.50, 56.88, 72.06, 121.44, 97.00, 42.25, 71.88, 70.63, 35.81, 83.19, 43.63, 40.06.

Calculate the average closing price.

Answer: $\bar{x} =$

The Arithmetic Mean: Example

The following is a list of closing stock prices:

31.69, 56.69, 65.50, 83.50, 56.88, 72.06, 121.44, 97.00, 42.25, 71.88, 70.63, 35.81, 83.19, 43.63, 40.06.

Calculate the average closing price.

Answer: $\bar{x} = \frac{972.21}{15} =$

The Arithmetic Mean: Example

The following is a list of closing stock prices:

31.69, 56.69, 65.50, 83.50, 56.88, 72.06, 121.44, 97.00, 42.25, 71.88, 70.63, 35.81, 83.19, 43.63, 40.06.

Calculate the average closing price.

Answer: $\bar{x} = \frac{972.21}{15} = 64.814$

Measures of Central Tendency: The Weighted Average

Often, some observations are more important than others. In this case, we assign a weight to each observation and calculate the weighted mean.

We denote the weighted average as $\bar{x}_w$ or $\bar{\mu}_w$. To calculate the weighted average, we multiply each observation x_i by its weight w_i, sum these up, and divide by the sum of weights.

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i \times x_i}{\sum\limits_{i=1}^{n} w_i}$$

The Weighted Average: Example 1

Suppose you score as follows in this course, calculate your average score with and without the weights.

	Exam 1	Exam 2	Exam 3	Participation	Pop Quizzes
Score	96%	91%	84%	94%	90%
Weight	20%	25%	50%	5%	5%

Answer:

$$\bar{x} = \frac{455}{5} = 91$$

$$\bar{x}_w = \frac{20 * 96 + 25 * 91 + 50 * 84 + 5 * 94 + 5 * 90}{20 + 25 + 50 + 5 + 5} = \frac{9315}{105} = 88.7$$

The Weighted Average: An example

Table below shows US census data on population and household income in NYC boroughs. Using this data we can calculate the two measures of HH income.

Borough	Household Income	Households
Bronx	34,284	480,323
Manhattan	71,656	745,089
Brooklyn	46,958	925,371
Queens	57,210	780,069

$$\bar{x} = \frac{34,284 + 71,656 + 46,958 + 57,210}{4} = 52527$$

$$\bar{x}_w = \frac{34,284 \times 480,323 + 71,656 \times 745,089 + 46,958 \times 925,371 + 57,210 \times 780,069}{480,323 + 745,089 + 925,371 + 780,069}$$

$$\bar{x}_w = 53888$$

The Weighted Mean: Example

A utility company spent approximately \$50,845, \$43,690, \$47,098, \$56,121, and \$49,369 on employee training for the years 2005 through 2009. Employees at the end of each year numbered 4738, 4637, 4540, 4397, and 4026, respectively. Using the annual number of employees as weights, what is the weighted mean for annual expenditures per employee during this period?

The Weighted Mean: Example

A utility company spent approximately \$50,845, \$43,690, \$47,098, \$56,121, and \$49,369 on employee training for the years 2005 through 2009. Employees at the end of each year numbered 4738, 4637, 4540, 4397, and 4026, respectively. Using the annual number of employees as weights, what is the weighted mean for annual expenditures per employee during this period?

Answer:

x	w	$x_i * w_i$
50845	4738	240903610
43690	4637	202590530
47098	4540	213824920
56121	4397	246764037
49369	4026	198759594
$\sum x_i$ 247123	$\sum w_i$ 22338	$\sum x_i * w_i$ 1102842691
$\bar{x} = 49424.6$	$\bar{x}_w = 49370.69975$	

Measures of Central Tendency: The Median

The median is the value separating the higher and lower halves of ordered observations in a sample or a population.

Example: The median test score of five students who scored 90, 80, 70, 95, and 75 is 80.

When the number of observations is even, the median becomes halfway between the two values in the middle when the data are arranged in order of size.

Example: The median net worth of our 2016 election candidates is $(10 + 22)/2 = \$16$ Millions.

One advantage of the median is that it is not affected by outliers. The median estimate of 16 million dollars is a much more accurate description of the typical candidate's net wroth than the 392.6 million dollars we got when we used the mean.

The Median: Example

The number of visits to a website is as follows:

65, 36, 52, 70, 37, 55, 63, 59, 68, 56, 63, 63, 43, 46, 73, 41, 47, 75, 75, and 54.

Determine the median number of visits.

Answer:

The Median: Example

The number of visits to a website is as follows:

65, 36, 52, 70, 37, 55, 63, 59, 68, 56, 63, 63, 43, 46, 73, 41, 47, 75, 75, and 54.

Determine the median number of visits.

Answer: First, you need to sort and count the observations

The Median: Example

The number of visits to a website is as follows:

65, 36, 52, 70, 37, 55, 63, 59, 68, 56, 63, 63, 43, 46, 73, 41, 47, 75, 75, and 54.

Determine the median number of visits.

Answer: First, you need to sort and count the observations :
36, 37, 41, 43, 46, 47, 52, 54, 55, 56, 59, 63, 63, 63, 65, 68, 70, 73, 75, 75.

The Median: Example

The number of visits to a website is as follows:

65, 36, 52, 70, 37, 55, 63, 59, 68, 56, 63, 63, 43, 46, 73, 41, 47, 75, 75, and 54.

Determine the median number of visits.

Answer: First, you need to sort and count the observations :
36, 37, 41, 43, 46, 47, 52, 54, 55, 56, 59, 63, 63, 63, 65, 68, 70, 73, 75, 75.

Since, $n = 20$, the median number of visits is

The Median: Example

The number of visits to a website is as follows:

65, 36, 52, 70, 37, 55, 63, 59, 68, 56, 63, 63, 43, 46, 73, 41, 47, 75, 75, and 54.

Determine the median number of visits.

Answer: First, you need to sort and count the observations :
36, 37, 41, 43, 46, 47, 52, 54, 55, 56, 59, 63, 63, 63, 65, 68, 70, 73, 75, 75.

Since, $n = 20$, the median number of visits is $\frac{56+59}{2} = 57.5$

Measures of Central Tendency: The Mode

The mode is the observation that occurs with the greatest frequency in a sample or a population.

Example: In the English monarchs data below, the mode age at accession is 22.

1 , 9 , 9 , 10 , 10 , 12 , 12 , 14 , 15 , 16 , 18 , 18 , 18 , 19 , 19 , 22 , 22 , 22 , 22 , 23 , 23 , 24 , 25 , 26 , 26 , 26 , 29 , 29 , 30 , 31 , 31 , 32 , 32 , 32 , 33 , 34 , 37 , 37 , 37 , 38 , 38 , 38 , 39 , 41 , 42 , 44 , 45 , 46 , 51 , 54 , 58 , 59 , 65

The Mode: Example

The following are scores on an exam:

80, 89, 90, 84, 84, 85, 83, 94, 70, 84, 59, 70, 84, 58, 68, 85, 70, 75, 55, 60, 65.

What is the mode score on this exam?

Answer: The mode is

The Mode: Example

The following are scores on an exam:

80, 89, 90, 84, 84, 85, 83, 94, 70, 84, 59, 70, 84, 58, 68, 85, 70, 75, 55, 60, 65.

What is the mode score on this exam?

Answer: The mode is 84

Comparing Measures of Central Tendency

Mean	Median	Mode
Gives equal consideration to all observations	Closely focuses on the center of the data	There may be more than one mode
Can be strongly influenced by outliers		Can be useful for categorical data

Measures of Dispersion: The Range

Definition

The ***range*** is the distance between the highest and lowest observations in the data.

Although the range is easy to use and understand, its weakness is that it is extremely sensitive to outliers.

Example

For our 2016 election candidates, the range of candidate networth is $4500 - 0.7 = 4499.3$ Millions. Without Trump, it is $58 - 0.7 = 57.3$ Millions.

Quantiles

Quantiles separate the data into equal-size groups after observations are arranged in order of numerical value.

Definitions

Percentiles divide data values into 100 parts of equal size, each comprising 1% of the observations. The median describes the 50^{th} percentile.

Deciles divide data values into 10 parts of equal size, each comprising 10% of the observations. The median is the 5^{th} decile.

Quartiles divide data values into four parts of equal size, each comprising 25% of the observations. The median describes the 2^{nd} quartile, below which 50% of the values fall.

Calculating Quartiles

Quartiles are calculated in a way very similarly to the way we calculate the median. Assuming the data are arranged from smallest to largest, the first, second, and third quartiles are calculates as follows:

$$
\begin{aligned}
Q_1 &= \text{Data value at position } \frac{n+1}{4} \\
Q_2 &= \text{Data value at position } \frac{2(n+1)}{4} = \text{The Median} \\
Q_3 &= \text{Data value at position } \frac{3(n+1)}{4}
\end{aligned}
$$

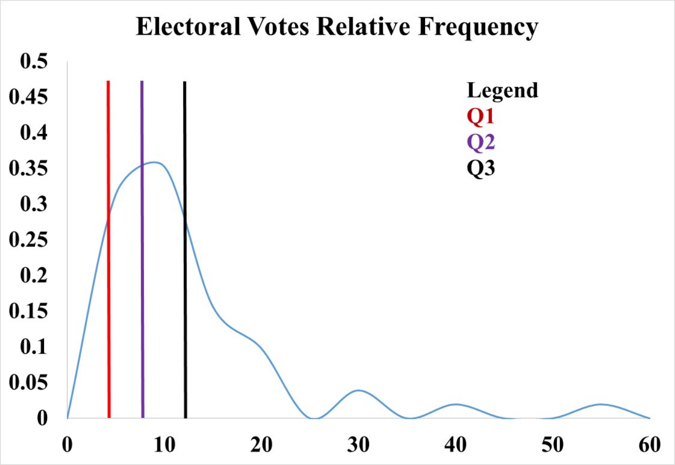
Where n is the number of observations in the data.

Calculating Quartiles: Electoral votes in 51 US States

State	Votes	State	Votes	State	Votes	State	Votes	State	Votes
AK	3	NH	4	CT	7	WI	10	OH	18
DE	3	RI	4	OK	7	AZ	11	IL	20
DC	3	NE	5	OR	7	IN	11	PA	20
MT	3	NM	5	KY	8	MA	11	FL	29
ND	3	WV	5	LA	8	TN	11	NY	29
SD	3	AR	6	AL	9	WA	12	TX	38
VT	3	IA	6	CO	9	VA	13	CA	55
WY	3	KS	6	SC	9	NJ	14		
HI	4	MS	6	MD	10	NC	15		
ID	4	NV	6	MN	10	GA	16		
ME	4	UT	6	MO	10	MI	16		

$Q_1 = \frac{51+1}{4} = 13^{th}$ Obs. $= 4$; $Q_2 = \frac{2(51+1)}{4} = 26^{th}$ Obs. $= 8$
$Q_3 = \frac{3(51+1)}{4} = 39^{th}$ Obs. $= 12$

Visualizing Quartiles



Calculating Quartiles

Calculate the 4 quartiles of following 20 exam scores: 56 , 57 , 58 , 60 , 61 , 64 , 65 , 70 , 70 , 72 , 74 , 76 , 76 , 78 , 80 , 81 , 82 , 85 , 91 , 94

Calculating Quartiles

Calculate the 4 quartiles of following 20 exam scores: 56 , 57 , 58 , 60 , 61 , 64 , 65 , 70 , 70 , 72 , 74 , 76 , 76 , 78 , 80 , 81 , 82 , 85 , 91 , 94

Answer

$Q_1 =$

$Q_2 =$

$Q_3 =$

Calculate the 4 quartiles of following 20 exam scores: 56 , 57 , 58 , 60 , 61 , 64 , 65 , 70 , 70 , 72 , 74 , 76 , 76 , 78 , 80 , 81 , 82 , 85 , 91 , 94

Answer

$$Q_1 = \frac{20+1}{4} = 5.25^{th} \text{ Obs.} \quad = 61 + 0.25 * (64 - 61) = 61.75$$

$$Q_2 =$$

$$Q_3 =$$

Calculating Quartiles

Calculate the 4 quartiles of following 20 exam scores: 56 , 57 , 58 , 60 , 61 , 64 , 65 , 70 , 70 , 72 , 74 , 76 , 76 , 78 , 80 , 81 , 82 , 85 , 91 , 94

Answer

$$Q_1 = \frac{20+1}{4} = 5.25^{th} \text{ Obs. } = 61 + 0.25 * (64 - 61) = 61.75$$

$$Q_2 = \frac{2(20+1)}{4} = 10.5^{th} \text{ Obs. } = 72 + 0.5 * (74 - 72) = 73$$

$$Q_3 =$$

Calculating Quartiles

Calculate the 4 quartiles of following 20 exam scores: 56 , 57 , 58 , 60 , 61 , 64 , 65 , 70 , 70 , 72 , 74 , 76 , 76 , 78 , 80 , 81 , 82 , 85 , 91 , 94

Answer

$$Q_1 = \frac{20+1}{4} = 5.25^{th} \text{ Obs.} \;\; = 61 + 0.25 * (64 - 61) = 61.75$$

$$Q_2 = \frac{2(20+1)}{4} = 10.5^{th} \text{ Obs.} \;\; = 72 + 0.5 * (74 - 72) = 73$$

$$Q_3 = \frac{3(20+1)}{4} = 15.75^{th} \text{ Obs.} \;\; = 80 + 0.75 * (81 - 80) = 80.75$$

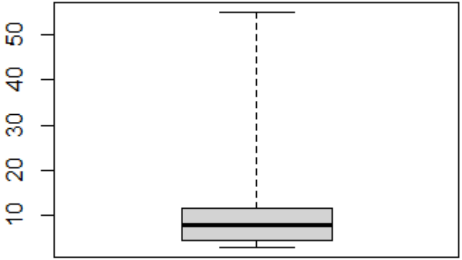
The Box-and-Whisker Plot (Box Plot)

Definition

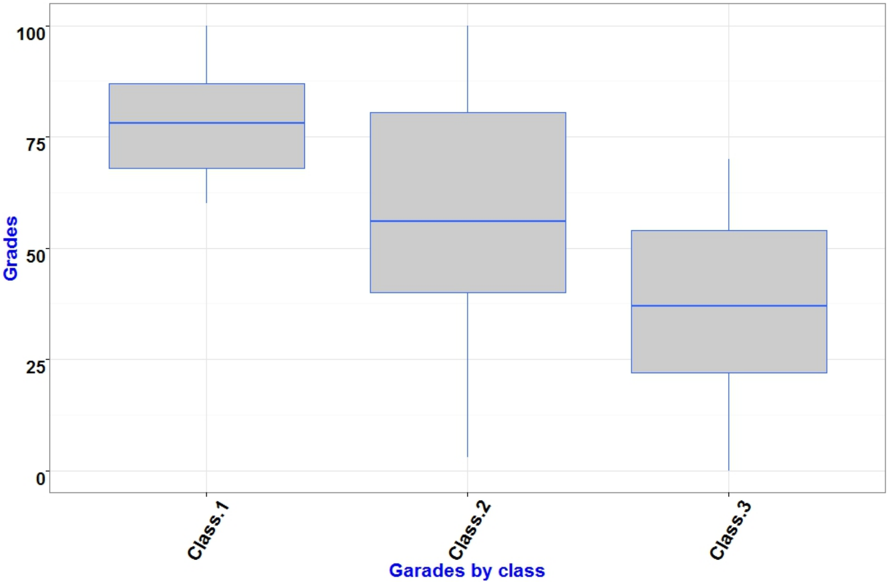
The ***box-and-whisker plot*** is a graphical device that simultaneously displays the first and third quartiles, the median, and the extreme values in the data, allowing us to easily identify these descriptors. We can also see whether the distribution is symmetrical or whether it is skewed either negatively or positively.

Box Plot: Electoral votes in 51 US States

$Q_1 = \frac{51+1}{4} = 13^{th}$ Obs. $= 4$; $Q_2 = \frac{2(51+1)}{4} = 26^{th}$ Obs. $= 8$
$Q_3 = \frac{3(51+1)}{4} = 39^{th}$ Obs. $= 12$



Box Plot: Example 1



Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \quad Q_3^{4cyl} = 30.4$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$$
$$Q_2^{4cyl} = 26 \quad Q_3^{4cyl} = 30.4$$
$$Max^{4cyl} = 33.9$$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \quad Q_3^{4cyl} = 30.4$

$Max^{4cyl} = 33.9$

$Min^{6cyl} = 17.8$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \qquad Q_3^{4cyl} = 30.4$

$Max^{4cyl} = 33.9$

$Min^{6cyl} = 17.8 \quad Q_1^{6cyl} = 18.1$

Cars Mileage by Number of Cylinders

mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \qquad Q_3^{4cyl} = 30.4$

$Max^{4cyl} = 33.9$

$Min^{6cyl} = 17.8 \quad Q_1^{6cyl} = 18.1$

$Q_2^{6cyl} = 19.7$

Cars Mileage by Number of Cylinders

mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \quad Q_3^{4cyl} = 30.4$

$Max^{4cyl} = 33.9$

$Min^{6cyl} = 17.8 \quad Q_1^{6cyl} = 18.1$

$Q_2^{6cyl} = 19.7 \quad Q_3^{6cyl} = 21$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \quad Q_3^{4cyl} = 30.4$

$Max^{4cyl} = 33.9$

$Min^{6cyl} = 17.8 \quad Q_1^{6cyl} = 18.1$

$Q_2^{6cyl} = 19.7 \quad Q_3^{6cyl} = 21$

$Max^{6cyl} = 21.4$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \quad Q_3^{4cyl} = 30.4$

$Max^{4cyl} = 33.9$

$Min^{6cyl} = 17.8 \quad Q_1^{6cyl} = 18.1$

$Q_2^{6cyl} = 19.7 \quad Q_3^{6cyl} = 21$

$Max^{6cyl} = 21.4$

$Min^{8cyl} = 10.4$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \qquad Q_3^{4cyl} = 30.4$

$Max^{4cyl} = 33.9$

$Min^{6cyl} = 17.8 \quad Q_1^{6cyl} = 18.1$

$Q_2^{6cyl} = 19.7 \qquad Q_3^{6cyl} = 21$

$Max^{6cyl} = 21.4$

$Min^{8cyl} = 10.4 \quad Q_1^{8cyl} = 14.05$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \quad Q_3^{4cyl} = 30.4$

$Max^{4cyl} = 33.9$

$Min^{6cyl} = 17.8 \quad Q_1^{6cyl} = 18.1$

$Q_2^{6cyl} = 19.7 \quad Q_3^{6cyl} = 21$

$Max^{6cyl} = 21.4$

$Min^{8cyl} = 10.4 \quad Q_1^{8cyl} = 14.05$

$Q_2^{8cyl} = 15.2$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$

$Q_2^{4cyl} = 26 \quad Q_3^{4cyl} = 30.4$

$Max^{4cyl} = 33.9$

$Min^{6cyl} = 17.8 \quad Q_1^{6cyl} = 18.1$

$Q_2^{6cyl} = 19.7 \quad Q_3^{6cyl} = 21$

$Max^{6cyl} = 21.4$

$Min^{8cyl} = 10.4 \quad Q_1^{8cyl} = 14.05$

$Q_2^{8cyl} = 15.2 \quad Q_3^{8cyl} = 16.625$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2

Use the data on mileage (mpg) to draw three box-plots by cylinder (cyl).

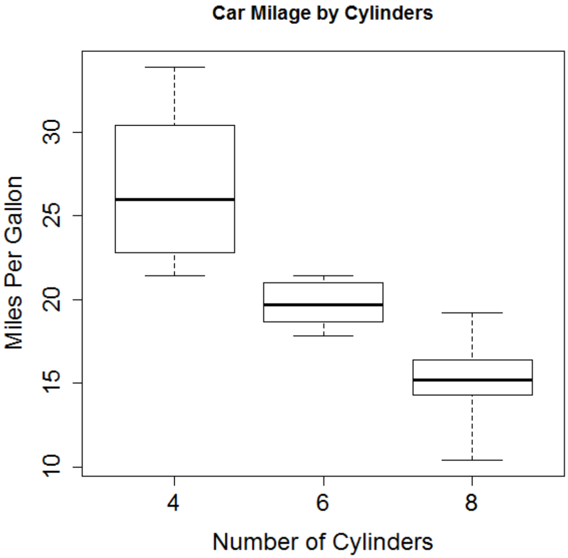
$$Min^{4cyl} = 21.4 \quad Q_1^{4cyl} = 22.8$$
$$Q_2^{4cyl} = 26 \quad Q_3^{4cyl} = 30.4$$
$$Max^{4cyl} = 33.9$$

$$Min^{6cyl} = 17.8 \quad Q_1^{6cyl} = 18.1$$
$$Q_2^{6cyl} = 19.7 \quad Q_3^{6cyl} = 21$$
$$Max^{6cyl} = 21.4$$

$$Min^{8cyl} = 10.4 \quad Q_1^{8cyl} = 14.05$$
$$Q_2^{8cyl} = 15.2 \quad Q_3^{8cyl} = 16.625$$
$$Max^{8cyl} = 19.2$$

Cars Mileage by Number of Cylinders		
mpg, cyl=4	mpg, cyl=6	mpg, cyl=8
22.8	21	18.7
24.4	21	14.3
22.8	21.4	16.4
32.4	18.1	17.3
30.4	19.2	15.2
33.9	17.8	10.4
21.5	19.7	10.4
27.3		14.7
26		15.5
30.4		15.2
21.4		13.3
		19.2
		15.8
		15

Box Plot: Example 2



Other Quartile Measures

Definition

The ***Interquartile Range*** is the difference between the third quartile and the first quartile, or $Q_3 - Q_1$. In percentile terms, this is the distance between the 75% and 25% values.

The ***Quartile Deviation*** is one-half the interquartile range, or $\frac{Q_3 - Q_1}{2}$

Example

For electoral votes data, $Q_1 = 4$, $Q_2 = 8$, and $Q_3 = 12$.

$$IR = Q_3 - Q_1 = 12 - 4 = 8$$
$$QD = \frac{Q_3 - Q_1}{2} = \frac{12 - 4}{2} = 4$$

Measures of Dispersion: The Mean Absolute Deviation

Definition

Residuals, or ***deviations***, are the differences between each data value in the set and the group mean: $x_i - \bar{x}$ or $x_i - \mu$

Definition

The mean absolute deviation (MAD) is found by summing up the absolute values of all residuals and dividing by the number of values in the set:

$$MAD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7		
R. Paul	2		
Christie	3		
Cruz	3.5		
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	
R. Paul	2		
Christie	3		
Cruz	3.5		
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x}\rvert$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3		
Cruz	3.5		
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5		
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\|x_i - \bar{x}\|$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\|x_i - \bar{x}\|$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$		

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x}\rvert$
Sanders	0.7	-18.5	
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	10.2
Kasich	10	-9.2	
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	10.2
Kasich	10	-9.2	9.2
J. Bush	22	2.8	
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	10.2
Kasich	10	-9.2	9.2
J. Bush	22	2.8	2.8
Carson	26	6.8	
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	10.2
Kasich	10	-9.2	9.2
J. Bush	22	2.8	2.8
Carson	26	6.8	6.8
Chafee	32	12.8	
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	10.2
Kasich	10	-9.2	9.2
J. Bush	22	2.8	2.8
Carson	26	6.8	6.8
Chafee	32	12.8	12.8
H. Clinton	45	25.8	
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x}\rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	10.2
Kasich	10	-9.2	9.2
J. Bush	22	2.8	2.8
Carson	26	6.8	6.8
Chafee	32	12.8	12.8
H. Clinton	45	25.8	25.8
Fiorina	58	38.8	
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x}\rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	10.2
Kasich	10	-9.2	9.2
J. Bush	22	2.8	2.8
Carson	26	6.8	6.8
Chafee	32	12.8	12.8
H. Clinton	45	25.8	25.8
Fiorina	58	38.8	38.8
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	10.2
Kasich	10	-9.2	9.2
J. Bush	22	2.8	2.8
Carson	26	6.8	6.8
Chafee	32	12.8	12.8
H. Clinton	45	25.8	25.8
Fiorina	58	38.8	38.8
	$\bar{x} = 19.2$	$\sum (x_i - \bar{x}) = 0$	$\sum \lvert x_i - \bar{x} \rvert = 174$

Calculating MAD networth for 2016 candidates

Candidate	Net Worth	Deviations $x_i - \bar{x}$	Absolute Deviations $\lvert x_i - \bar{x} \rvert$
Sanders	0.7	-18.5	18.5
R. Paul	2	-17.2	17.2
Christie	3	-16.2	16.2
Cruz	3.5	-15.7	15.7
Huckabee	9	-10.2	10.2
Kasich	10	-9.2	9.2
J. Bush	22	2.8	2.8
Carson	26	6.8	6.8
Chafee	32	12.8	12.8
H. Clinton	45	25.8	25.8
Fiorina	58	38.8	38.8
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	$\sum \lvert x_i - \bar{x} \rvert = 174$

$$MAD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{174}{11} = 15.8$$

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	
77	
78	
81	
83	
87	
90	
93	
93	
94	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	
77	
78	
81	
83	
87	
90	
93	
93	
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	
78	
81	
83	
87	
90	
93	
93	
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	
81	
83	
87	
90	
93	
93	
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	
83	
87	
90	
93	
93	
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	4.1
83	
87	
90	
93	
93	
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	4.1
83	2.1
87	
90	
93	
93	
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	4.1
83	2.1
87	1.9
90	
93	
93	
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	4.1
83	2.1
87	1.9
90	4.9
93	
93	
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	4.1
83	2.1
87	1.9
90	4.9
93	7.9
93	
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	4.1
83	2.1
87	1.9
90	4.9
93	7.9
93	7.9
94	
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	4.1
83	2.1
87	1.9
90	4.9
93	7.9
93	7.9
94	8.9
$\sum_{i=1}^{n} x_i = 851$	
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	4.1
83	2.1
87	1.9
90	4.9
93	7.9
93	7.9
94	8.9
$\sum_{i=1}^{n} x_i = 851$	$\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert = 63$
$\bar{x} = 85.1$	

Calculating MAD: Example

Calculate the MAD of 75 , 77 , 78 , 81 , 83 , 87 , 90 , 93 , 93 , 94

x_i	$\lvert x_i - \bar{x} \rvert$
75	10.1
77	8.1
78	7.1
81	4.1
83	2.1
87	1.9
90	4.9
93	7.9
93	7.9
94	8.9
$\sum_{i=1}^{n} x_i = 851$	$\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert = 63$
$\bar{x} = 85.1$	$MAD = 6.3$

Measures of Dispersion: The Variance

Definition

For a population, the variance, σ^2, "sigma squared", is the average of squared differences between the N data values and the mean μ. For a sample, the variance, s^2, is the sum of the squared differences between the n data values and the mean, $\bar{x}$, divided by $(n-1)$.

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7		
R. Paul	2		
Christie	3		
Cruz	3.5		
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7		
R. Paul	2		
Christie	3		
Cruz	3.5		
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2		
Christie	3		
Cruz	3.5		
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3		
Cruz	3.5		
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5		
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9		
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10		
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10	-9.2	84.64
J. Bush	22		
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10	-9.2	84.64
J. Bush	22	2.8	7.84
Carson	26		
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10	-9.2	84.64
J. Bush	22	2.8	7.84
Carson	26	6.8	46.24
Chafee	32		
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10	-9.2	84.64
J. Bush	22	2.8	7.84
Carson	26	6.8	46.24
Chafee	32	12.8	163.84
H. Clinton	45		
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10	-9.2	84.64
J. Bush	22	2.8	7.84
Carson	26	6.8	46.24
Chafee	32	12.8	163.84
H. Clinton	45	25.8	665.64
Fiorina	58		
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10	-9.2	84.64
J. Bush	22	2.8	7.84
Carson	26	6.8	46.24
Chafee	32	12.8	163.84
H. Clinton	45	25.8	665.64
Fiorina	58	38.8	1505.44
	$\bar{x} = 19.2$		

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10	-9.2	84.64
J. Bush	22	2.8	7.84
Carson	26	6.8	46.24
Chafee	32	12.8	163.84
H. Clinton	45	25.8	665.64
Fiorina	58	38.8	1505.44
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10	-9.2	84.64
J. Bush	22	2.8	7.84
Carson	26	6.8	46.24
Chafee	32	12.8	163.84
H. Clinton	45	25.8	665.64
Fiorina	58	38.8	1505.44
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	$\sum(x_i - \bar{x})^2 = 3724.7$

Calculating variance: candidates net worth

Candidate	Net Worth	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
Sanders	0.7	-18.5	342.25
R. Paul	2	-17.2	295.84
Christie	3	-16.2	262.44
Cruz	3.5	-15.7	246.49
Huckabee	9	-10.2	104.04
Kasich	10	-9.2	84.64
J. Bush	22	2.8	7.84
Carson	26	6.8	46.24
Chafee	32	12.8	163.84
H. Clinton	45	25.8	665.64
Fiorina	58	38.8	1505.44
	$\bar{x} = 19.2$	$\sum(x_i - \bar{x}) = 0$	$\sum(x_i - \bar{x})^2 = 3724.7$

$$s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1) = 3724.7/(11-1) = 372.47$$

Measures of Dispersion: The Standard Deviation

Definition

The ***standard deviation*** is the positive square root of the variance of either a population or a sample.

	Sample	Population
Standard Deviation	$\sqrt{s^2} = s$	$\sqrt{\sigma^2} = \sigma$

Example

The standard deviation of 2016 candidate networth is $s = \sqrt{s^2} = \sqrt{372.47} = 19.29$

Calculating variance & Standard deviation : Example

Suppose the seven most visited shopping websites had the following visits in millions: eBay (62), Amazon (41), Wal-Mart (24), Shopping.com (26), Target (22), Apple Computer (18), and Overstock.com (17). Considering these as a population, determine: the mean, variance, and standard deviation.

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62		
Amazon	41		
Wal-Mart	24		
Shopping.com	26		
Target	22		
Apple	18		
Overstock.com	17		

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62		
Amazon	41		
Wal-Mart	24		
Shopping.com	26		
Target	22		
Apple	18		
Overstock.com	17		
	$\mu = 30$		

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41		
Wal-Mart	24		
Shopping.com	26		
Target	22		
Apple	18		
Overstock.com	17		
	$\mu = 30$		

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24		
Shopping.com	26		
Target	22		
Apple	18		
Overstock.com	17		
	$\mu = 30$		

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24	-6	36
Shopping.com	26		
Target	22		
Apple	18		
Overstock.com	17		
	$\mu = 30$		

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24	-6	36
Shopping.com	26	-4	16
Target	22		
Apple	18		
Overstock.com	17		
	$\mu = 30$		

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24	-6	36
Shopping.com	26	-4	16
Target	22	-8	64
Apple	18		
Overstock.com	17		
	$\mu = 30$		

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24	-6	36
Shopping.com	26	-4	16
Target	22	-8	64
Apple	18	-12	144
Overstock.com	17		
	$\mu = 30$		

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24	-6	36
Shopping.com	26	-4	16
Target	22	-8	64
Apple	18	-12	144
Overstock.com	17	-13	169
	$\mu = 30$		

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24	-6	36
Shopping.com	26	-4	16
Target	22	-8	64
Apple	18	-12	144
Overstock.com	17	-13	169
	$\mu = 30$	$\sum = 0$	

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24	-6	36
Shopping.com	26	-4	16
Target	22	-8	64
Apple	18	-12	144
Overstock.com	17	-13	169
	$\mu = 30$	$\sum = 0$	$\sum = 1574$

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24	-6	36
Shopping.com	26	-4	16
Target	22	-8	64
Apple	18	-12	144
Overstock.com	17	-13	169
	$\mu = 30$	$\sum = 0$	$\sum = 1574$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} = \frac{1574}{7} = 224.857$$

Calculating variance & Standard deviation : Example

Website	Visits (x_i)	$x_i - \mu$	$(x_i - \mu)^2$
eBay	62	32	1024
Amazon	41	11	121
Wal-Mart	24	-6	36
Shopping.com	26	-4	16
Target	22	-8	64
Apple	18	-12	144
Overstock.com	17	-13	169
	$\mu = 30$	$\sum = 0$	$\sum = 1574$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} = \frac{1574}{7} = 224.857$$
$$\sigma = \sqrt{\sigma^2} = \sqrt{224.857} = 14.995$$

The variance and standard deviation

- Unless all observations have same value, the variance and standard deviation cant be zero.
- If the same number is added to or subtracted from all the values, the variance and standard deviation remain unchanged.
- Standard deviation based on (N) divisor = Standard deviation based on (N-1) divisor $\times\sqrt{\frac{N-1}{N}}$

Understanding Distributions

> "My grades are like lightning. They are liable to strike anywhere"
>
> — Thorstein Veblen

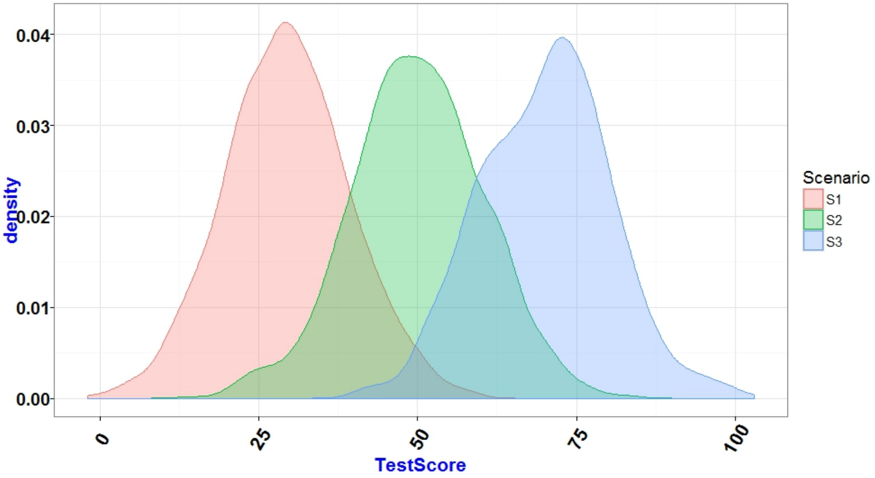
Suppose a 1000 students take a test. The test grades can be distributed in many ways. The shape of grades distribution determines the relative values of the mean, median, and mode.

The distributions can be ***symmetric*** or ***skewed***. Lets explore these concepts using the possible scenarios the test results can follow.

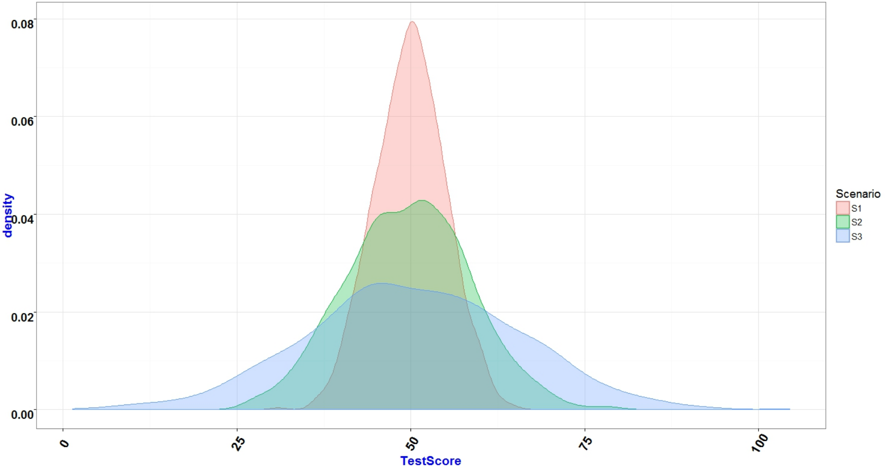
Definitions

- If mean= median = mode, then the shape of the distribution is ***symmetric***.
- If mode < median < mean, the the shape of the distribution trails to the right, and the distribution is ***positively skewed***.
- If mean < median < mode, then the shape of the distribution trails to the left, and the distribution is ***negatively skewed***.

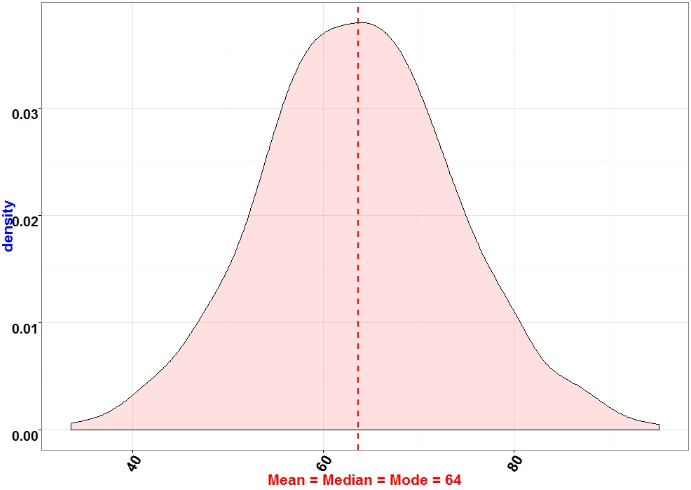
Symmetric Distributions: Different means, same SDV



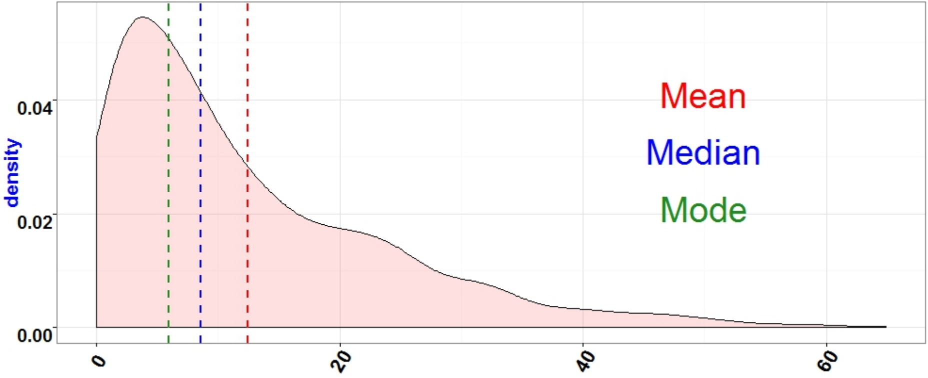
Symmetric Distributions: Same mean, different SDVs



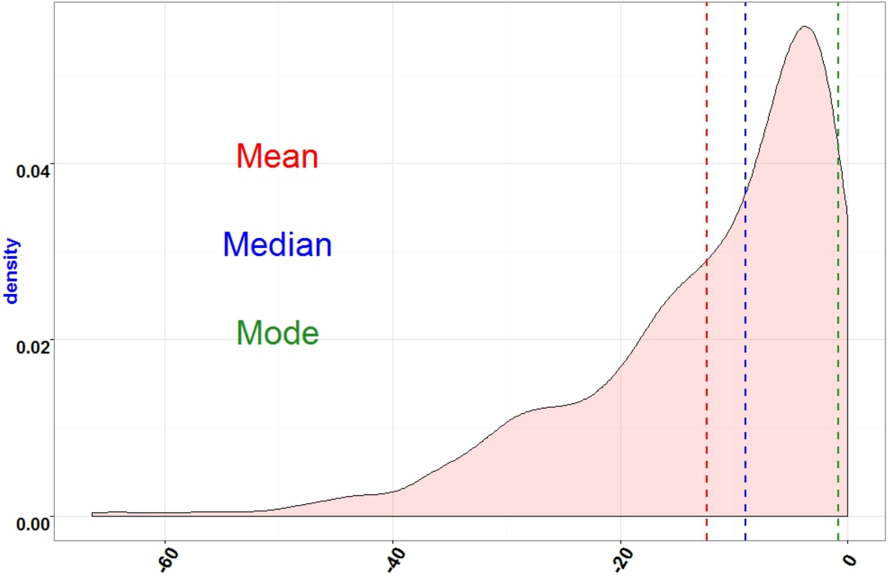
The Mean, Median, and Mode in Symmetric Distributions



Positively skewed distributions



Negatively skewed distributions



Chebyshev's Theorem

With a small standard deviation, individual observations will tend to be closer to the mean. Conversely, If the standard deviation is large, individual observations will be scattered widely about their mean.

Theorem (Chebyshev's Theorem)

For either a sample or a population, the percentage of observations that fall within k standard deviations of the mean will be at least

$$\left(1 - \frac{1}{k^2}\right) \times 100$$

for all $k > 1$.

Chebyshev's Theorem: Example

Bellow are ages of most 2016 presidential candidates, with $\bar{x} = 60.48$ and $s = 9.23$:

45 , 45 , 46 , 49 , 54 , 54 , 54 , 58 , 61 , 61 , 62 , 63 , 64 , 64 , 65 , 66 , 69 , 70 , 71 , 74 , 75.

Using Chebyshev's theorem, what proportion of the data falls within 1.2 and 2 standard deviations from the mean, respectively?

Answer:

Chebyshev's Theorem: Example

Bellow are ages of most 2016 presidential candidates, with $\bar{x} = 60.48$ and $s = 9.23$:
45 , 45 , 46 , 49 , 54 , 54 , 54 , 58 , 61 , 61 , 62 , 63 , 64 , 64 , 65 , 66 , 69 , 70 , 71 , 74 , 75.
Using Chebyshev's theorem, what proportion of the data falls within 1.2 and 2 standard deviations from the mean, respectively?
Answer: Using Chebyshev's theorem, if $k = 1.2$, then at least

$(1 - \frac{1}{1.2^2})100 = (1 - \frac{1}{1.44})100 = 30.56\%$ of the observations fall within $60.8 \pm 1.2 \times 9.23 = [49.72, 71.88]$.

Chebyshev's Theorem: Example

Bellow are ages of most 2016 presidential candidates, with $\bar{x} = 60.48$ and $s = 9.23$:
45 , 45 , 46 , 49 , 54 , 54 , 54 , 58 , 61 , 61 , 62 , 63 , 64 , 64 , 65 , 66 , 69 , 70 , 71 , 74 , 75.
Using Chebyshev's theorem, what proportion of the data falls within 1.2 and 2 standard deviations from the mean, respectively?
Answer: Using Chebyshev's theorem, if $k = 1.2$, then at least

$(1 - \frac{1}{1.2^2})100 = (1 - \frac{1}{1.44})100 = 30.56\%$ of the observations fall within $60.8 \pm 1.2 \times 9.23 = [49.72, 71.88]$.

The percentage of data values falling within the interval $[49.72, 71.88]$ is actually a 76.19%, well above the minimum of 30.56% predicted by the theorem.

Chebyshev's Theorem: Example, Contd.

Bellow are ages of most 2016 presidential candidates, with $\bar{x} = 60.48$ and $s = 9.23$:
45 , 45 , 46 , 49 , 54 , 54 , 54 , 58 , 61 , 61 , 62 , 63 , 64 , 64 , 65 , 66 , 69 , 70 , 71 , 74 , 75.
Using Chebyshev's theorem, what proportion of the data falls within 1.2 and 2 standard deviations from the mean, respectively?

Answer:

Chebyshev's Theorem: Example, Contd.

Bellow are ages of most 2016 presidential candidates, with $\bar{x} = 60.48$ and $s = 9.23$:
45 , 45 , 46 , 49 , 54 , 54 , 54 , 58 , 61 , 61 , 62 , 63 , 64 , 64 , 65 , 66 , 69 , 70 , 71 , 74 , 75.
Using Chebyshev's theorem, what proportion of the data falls within 1.2 and 2 standard deviations from the mean, respectively?

Answer: Using Chebyshev's theorem, if $k = 2$, then at least $(1 - \frac{1}{2^2})100 = (1 - \frac{1}{4})100 = 75\%$ of the observations fall within $60.8 \pm 2 \times 9.23 = [42, 78.94]$.

Chebyshev's Theorem: Example, Contd.

Bellow are ages of most 2016 presidential candidates, with $\bar{x} = 60.48$ and $s = 9.23$:
45 , 45 , 46 , 49 , 54 , 54 , 54 , 58 , 61 , 61 , 62 , 63 , 64 , 64 , 65 , 66 , 69 , 70 , 71 , 74 , 75.
Using Chebyshev's theorem, what proportion of the data falls within 1.2 and 2 standard deviations from the mean, respectively?

Answer: Using Chebyshev's theorem, if $k = 2$, then at least $(1 - \frac{1}{2^2})100 = (1 - \frac{1}{4})100 = 75\%$ of the observations fall within $60.8 \pm 2 \times 9.23 = [42, 78.94]$.

The percentage of data values falling within the interval $[42, 78.94]$ is actually a 100%, well above the minimum of 75% predicted by the theorem.

Chebyshev's Theorem: Example

According to Chebyshev's theorem, what percentage of observations should fall

- within 2.5 standard deviations of the mean?

Chebyshev's Theorem: Example

According to Chebyshev's theorem, what percentage of observations should fall

- within 2.5 standard deviations of the mean?
- * 84%
- within 3.0 standard deviations of the mean?

Chebyshev's Theorem: Example

According to Chebyshev's theorem, what percentage of observations should fall

- within 2.5 standard deviations of the mean?
* 84%
- within 3.0 standard deviations of the mean?
* 88.89%
- within 5.0 standard deviations of the mean?

Chebyshev's Theorem: Example

According to Chebyshev's theorem, what percentage of observations should fall

- within 2.5 standard deviations of the mean?

* 84%

- within 3.0 standard deviations of the mean?

* 88.89%

- within 5.0 standard deviations of the mean?

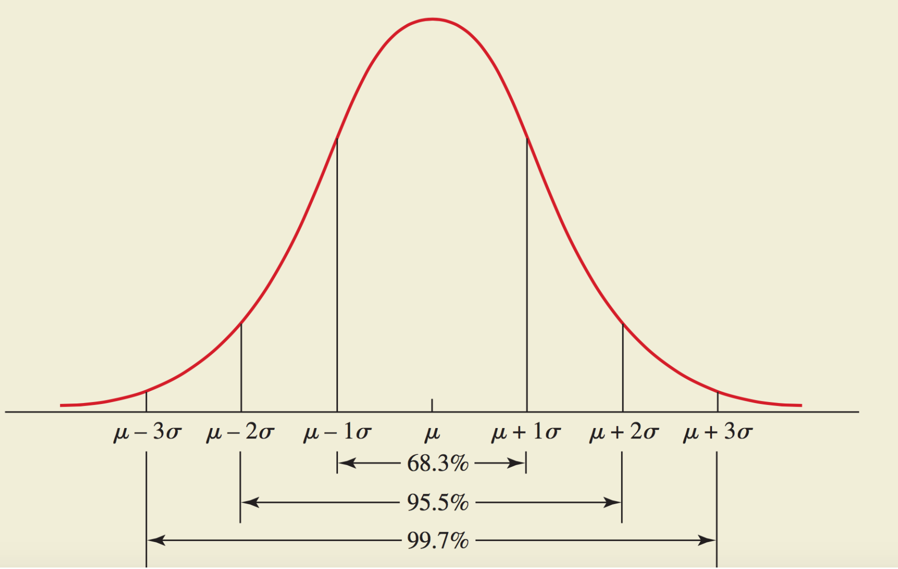
* 96%

The Empirical Rule

For distributions that are ***bell shaped and symmetrical*** :

- About 68% of the observations will fall within 1 standard deviation of the mean.
- About 95% of the observations will fall within 2 standard deviations of the mean.
- Practically all of the observations ($\approx 99.7\%$) will fall within 3 standard deviations of the mean.

The Empirical Rule



The Empirical Rule: Example 1

The manufacturer of an extended-life light bulb claims the bulb has an average life of 12,000 hours, with a standard deviation of 500 hours. If the distribution is bell shaped and symmetrical, what is the approximate percentage of these bulbs that will last

- between 11,000 and 13,000 hours?

The Empirical Rule: Example 1

The manufacturer of an extended-life light bulb claims the bulb has an average life of 12,000 hours, with a standard deviation of 500 hours. If the distribution is bell shaped and symmetrical, what is the approximate percentage of these bulbs that will last

- between 11,000 and 13,000 hours?
  * 95% (± 2 standard deviations of the mean).
- over 12,500 hours?

The Empirical Rule: Example 1

The manufacturer of an extended-life light bulb claims the bulb has an average life of 12,000 hours, with a standard deviation of 500 hours. If the distribution is bell shaped and symmetrical, what is the approximate percentage of these bulbs that will last

- between 11,000 and 13,000 hours?
  * 95% (± 2 standard deviations of the mean).
- over 12,500 hours?
  * 16%, or $50\% - 34\%$. (± 1 standard deviations of the mean).
- less than 11,000 hours?

The Empirical Rule: Example 1

The manufacturer of an extended-life light bulb claims the bulb has an average life of 12,000 hours, with a standard deviation of 500 hours. If the distribution is bell shaped and symmetrical, what is the approximate percentage of these bulbs that will last

- between 11,000 and 13,000 hours?
  * 95% (± 2 standard deviations of the mean).
- over 12,500 hours?
  * 16%, or $50\% - 34\%$. (± 1 standard deviations of the mean).
- less than 11,000 hours?
  * 2.25%, or $(99.7 - 95.5)/2 + (100 - 99.7)/2$
- between 11,500 and 13,000 hours?

The Empirical Rule: Example 1

The manufacturer of an extended-life light bulb claims the bulb has an average life of 12,000 hours, with a standard deviation of 500 hours. If the distribution is bell shaped and symmetrical, what is the approximate percentage of these bulbs that will last

- between 11,000 and 13,000 hours?
  * 95% (± 2 standard deviations of the mean).
- over 12,500 hours?
  * 16%, or $50\% - 34\%$. (± 1 standard deviations of the mean).
- less than 11,000 hours?
  * 2.25%, or $(99.7 - 95.5)/2 + (100 - 99.7)/2$
- between 11,500 and 13,000 hours?
  * 81.5%, obtained by 34% (the area between the mean and 11,500) plus 47.5% (the area from the mean to 13,000).

The Empirical Rule: Example 2

IQ scores have a bell-shaped distribution with mean $\mu = 100$ and a standard deviation $\sigma = 10$, what percentage of IQ scores falls:

- above 110

The Empirical Rule: Example 2

IQ scores have a bell-shaped distribution with mean $\mu = 100$ and a standard deviation $\sigma = 10$, what percentage of IQ scores falls:

- above 110
  * $\approx 16\%$
- bellow 80

The Empirical Rule: Example 2

IQ scores have a bell-shaped distribution with mean $\mu = 100$ and a standard deviation $\sigma = 10$, what percentage of IQ scores falls:

- above 110
  * $\approx 16\%$
- bellow 80
  * $\approx 2.25\%$
- above 130

The Empirical Rule: Example 2

IQ scores have a bell-shaped distribution with mean $\mu = 100$ and a standard deviation $\sigma = 10$, what percentage of IQ scores falls:

- above 110
- * $\approx 16\%$
- bellow 80
- * $\approx 2.25\%$
- above 130
- * $\approx 0.15\%$
- between 80 and 90

The Empirical Rule: Example 2

IQ scores have a bell-shaped distribution with mean $\mu = 100$ and a standard deviation $\sigma = 10$, what percentage of IQ scores falls:

- above 110

  * $\approx 16\%$
- bellow 80

  * $\approx 2.25\%$
- above 130

  * $\approx 0.15\%$
- between 80 and 90

  * $\approx 13.5\%$

Standardized Data

Standardizing the data involves expressing each data value in terms of its distance (in standard deviations) from the mean. For each observation in a sample:

$$
\begin{aligned}
z_i = \frac{x_i - \bar{x}}{s} \text{where } z_i &= \text{standardized value for the } i^{th} \text{ observation} \\
\bar{x} &= \text{sample mean} \\
x_i &= \text{the } i^{th} \text{data value} \\
s &= \text{sample standard deviation}
\end{aligned}
$$

When data represent a population, μ replaces $\bar{x}$, and σ replaces s
A standardized value that is large (positive or negative) is relatively unusual in occurrence.
A standardized distribution will always have a mean of 0 and a standard deviation of 1.

Standardized Data: Heart rate (bpm)

bpm
35
40
50
60
65
70
75
80
90
100
$\Sigma = 665$
$\bar{x} = 66.50$

Standardized Data: Heart rate (bpm)

bpm	$(x_i - \bar{x})^2$
35	992.25
40	702.25
50	272.25
60	42.25
65	2.25
70	12.25
75	72.25
80	182.25
90	552.25
100	1122.25
$\Sigma = 665$	$\Sigma = 3952.50$
$\bar{x} = 66.50$	$s^2 = 439.17$
	$s = 20.96$

Standardized Data: Heart rate (bpm)

bpm	$(x_i - \bar{x})^2$	z_i
35	992.25	-1.50
40	702.25	-1.26
50	272.25	-0.79
60	42.25	-0.31
65	2.25	-0.07
70	12.25	0.17
75	72.25	0.41
80	182.25	0.64
90	552.25	1.12
100	1122.25	1.60
$\Sigma = 665$	$\Sigma = 3952.50$	$\Sigma = 0$
$\bar{x} = 66.50$	$s^2 = 439.17$	
	$s = 20.96$	

Standardized Data: Heart rate (bpm)

bpm	$(x_i - \bar{x})^2$	z_i	$(z_i - \bar{z})^2$
35	992.25	-1.50	2.26
40	702.25	-1.26	1.60
50	272.25	-0.79	0.62
60	42.25	-0.31	0.10
65	2.25	-0.07	0.01
70	12.25	0.17	0.03
75	72.25	0.41	0.16
80	182.25	0.64	0.41
90	552.25	1.12	1.26
100	1122.25	1.60	2.56
$\Sigma = 665$	$\Sigma = 3952.50$	$\Sigma = 0$	$\Sigma = 9$
$\bar{x} = 66.50$	$s^2 = 439.17$		$s^2 = 1$
	$s = 20.96$		$s = 1$

The Coefficient of Variation

Definition

Expressing the standard deviation as a percentage of the mean, the ***coefficient of variation (CV)*** indicates the relative amount of dispersion in the data. It is a good measure of volatility.

$$CV = \frac{\sigma}{\mu} \times 100, \text{ Or } CV = \frac{s}{\bar{x}} \times 100$$